

IN THE CLAIMS

This listing of the claim will replace all prior versions and listings of claim in the present application.

Listing of Claims

Claims 1 and 2 (canceled).

3. (currently amended) A similar document search method, implemented in a document search and retrieval system including a memory for storing a document database having documents and information concerning the documents and a processor for processing data of the document database, of searching to find and retrieve a document similar to a seeds document from said document database, the method, comprising the steps executed by the processor, of:

first extracting at least one characteristic word from the seeds document including desired retrieval contents;

if the at least one characteristic word extracted by the first extracting step is one of a compound characteristic phrase constructed by a plurality of constituent characteristic words included in the compound characteristic phrase, second extracting of characteristic words included in said compound characteristic phrase and each of said plurality of constituent characteristic words of said compound characteristic phrase;

calculating, according to the characteristic words extracted by the second extracting step, a degree of similarity between the seeds document and each of the document documents stored on the document database, by generating characteristic vectors using the characteristic words including the compound characteristic phrase and the constituent characteristic words by

which the compound characteristic phrase is constructed, the characteristic vectors being extracted from the seeds document and each of the documents stored on the document database; and

outputting a retrieval result, including an indication of the degree of similarity between each of the documents stored on the database and the seeds document, as a result of the degree of similarity between the seeds document and each of the documents stored on the document database as calculated by the similarity calculating step, thereby permitting documents having a high degree of similarity to the seeds document to be retrieved,

wherein the similarity calculating step includes the steps of:

calculating a weighting coefficient corresponding to a distance between said constituent characteristic words on the document stored on the document database, said distance being calculated based on term appearance positions of said constituent characteristic words in a phrase including said characteristic words extracted in said second extracting step on the document stored on the document database; and

calculating similarity using cosines of angles between the characteristic vectors extracted from the seeds document and the characteristic vectors extracted from each of the documents stored on the document database involving the weighting coefficient.

Claims 4 and 5 (canceled).

6. (currently amended) A similar document search system for searching to find and retrieve a document similar to a seeds document from a document database, the system, comprising:

a memory for storing the document database having documents and information concerning the documents; and

a processor for processing data of the document database,

wherein said processor comprises:

a document analyzer processor for extracting at least one characteristic word from the seeds document including desired retrieval contents,

a characteristic word extractor processor for, if the at least one characteristic word extracted by the document analyzer processor is one of a compound characteristic phrase constructed by a plurality of constituent characteristic words included in the compound characteristic phrase, extracting characteristic words included in said compound characteristic phrase and each of said plurality of constituent characteristic words of said compound characteristic phrase,

a seeds document similarity calculator processor for calculating, according to the characteristic words extracted by the characteristic word extractor processor, a degree of similarity between the seeds document and each of the document-documents stored on the document database, by generating characteristic vectors using the characteristic words including the compound characteristic phrase and the constituent characteristic words by which the compound characteristic phrase is constructed, the characteristic

vectors being extracted from the seeds document and each of the documents stored on the document database; and

a retrieval result output processor for outputting a retrieval result, including an indication of the degree of similarity between each of the documents stored on the database and the seeds document, as a result of the degree of similarity between the seeds document and each of the documents stored on the document database as calculated by the seeds document similarity calculator processor, thereby permitting documents having a high degree of similarity to the seeds document to be retrieved.

wherein the seeds document similarity calculator processor includes:

a weighting coefficient calculator processor for calculating a weighting coefficient corresponding to a distance between said constituent characteristic words on the document stored on the document database, said distance being calculated based on term appearance positions of said constituent characteristic words in a phrase including said characteristic words extracted in said second extracting step on the document stored on the document database, and

a calculator processor for calculating similarity using cosine of angle between the characteristic vectors extracted from the seeds document and the characteristic vectors extracted from the document stored on the document database involving the weighting coefficient.

Claims 7 and 8 (canceled).

9. (currently amended) A program, stored on a storage medium, executable by a document search and retrieval system including a memory for storing a document database having documents and information concerning the documents and a processor for processing data of the document database for searching to find and retrieve a document similar to a seeds document from said document database, the program, when executed cause the document search and retrieval system, to perform the steps of:

first extracting at least one characteristic word from the seeds document including desired retrieval contents;

if the at least one characteristic word extracted by the first extracting step is one of a compound characteristic phrase constructed by a plurality of constituent characteristic words included in the compound characteristic phrase, second extracting of characteristic words included in said compound characteristic phrase and each of said plurality of constituent characteristic words of said compound characteristic phrase;

calculating, according to the characteristic words extracted by the second extracting step, a degree of similarity between the seeds document and each of the document documents stored on the document database, by generating characteristic vectors using the characteristic words including the compound characteristic phrase and the constituent characteristic words by which the compound characteristic phrase is constructed, the characteristic vectors being extracted from the seeds document and each of the documents stored on the document database; and

outputting a retrieval result, including an indication of the degree of similarity between each of the documents stored on the database and the

seeds document, as a result of the degree of similarity between the seeds document and each of the documents stored on the document database as calculated by the calculating step, thereby permitting documents having a high degree of similarity to the seeds document to be retrieved,

wherein the calculating step includes the steps of:

calculating a weighting coefficient corresponding to a distance calculated by term appearance position on the seeds document between a constituent characteristic word and another constituent characteristic word which are extracted from one compound characteristic phrase, and

calculating similarity using cosine of angle between the characteristic vectors extracted from the seeds document and the characteristic vectors extracted from the document stored on the document database involving the weighting coefficient.